



# Adoption of Ontological Data Models by NASA Life Science Systems

Dan Berrios NASA Ames Research Center

ESIP Jan Mtg

January 2023

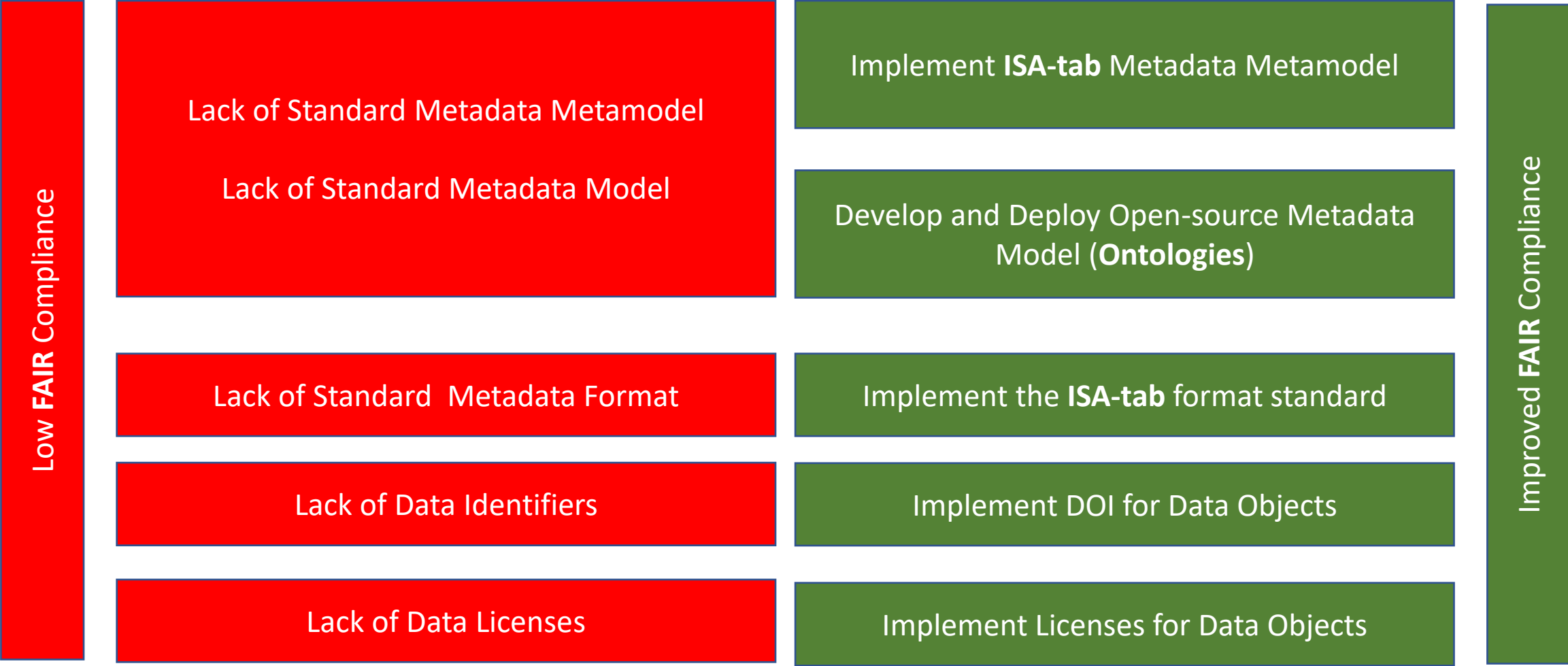
- FAIR and FAIR Metrics
- Guidance and Directives
- Assessing NGLSDA FAIR Compliance
- Implementing FAIR Compliance
  - Approaches
  - Prioritization
- FAIR Compliance Workbenches and Dashboards
- Summary



# FAIR Shortcomings of NASA LS Repositories



- Metadata harmonization and linked data are **key features** of **FAIR systems**
- Metadata harmonization and linked data are **foundations** of **Open Science**, in that they enhance transparency of investigations and facilitate scientific reproducibility.
- The Problem with NASA Life Sciences systems
  - Lack of use of community-based standards for metadata
  - Lack of uniformity of metadata values within and across (**metadata harmonization**)
  - Partially captured correspondences between metadata (**linked data**)
    - “This specimen in this experiment is a sample of that organism in that experiment”
    - “This instrument used in this experiment is the same as that instrument used in that experiment”





# Ontologies



- Ontologies are knowledge models, like taxonomies, but “suped up”.
- They let users classify things according to their attributes, but also support creating broader rules for things based on relationships between them.
- For example, through an ontology, you could declare a “Life Sciences Hardware” class with instances that can be “part of” one or more payload. This lets users define and refer to things by the way they are linked to other things, not just by where they appear in a taxonomy.
- Ontologies also support automated reasoning. The ontology can let you know if you create two rules for things which are in conflict with each other and why. This insures your model is “logical”
- Like taxonomies, queries can be expanded using the relationship knowledge captured in ontologies. “Find me experiments using hardware on vehicle payloads with attributes XYZ”



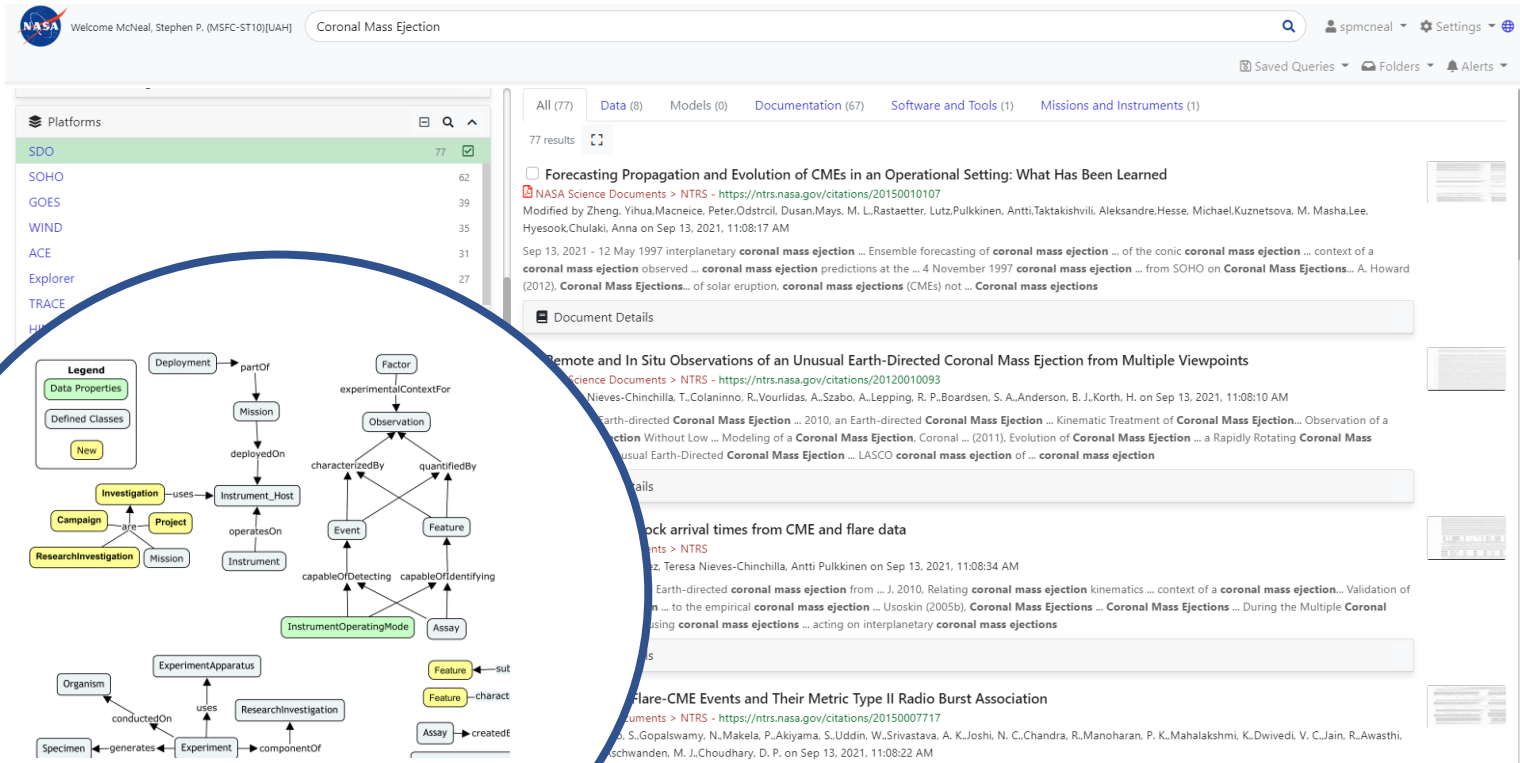
# Biomedical Investigation Ontologies



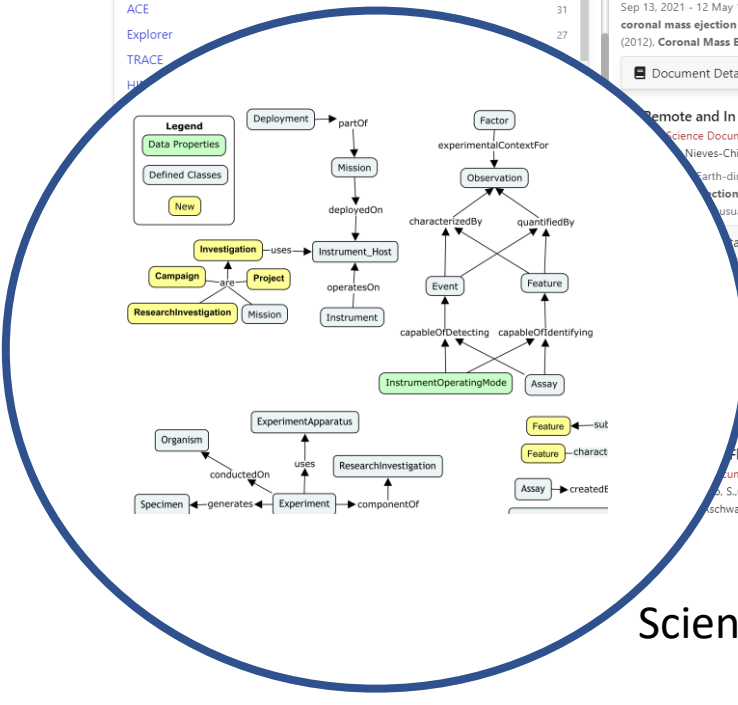
- OBO Foundry (~ 200 ontologies)
  - OBI Ontology for Biomedical Investigations
  - GO Gene Ontology
  - ENVO Environment Ontology
  - RBO Radiation Biology Ontology
- W3C
  - SOSA/SSN (Semantic Sensor Network)
  - TO Time Ontology
- NIH / NCBO (National Center for Biomedical Ontology) (1136 Ontologies, and counting)
  - NCBO Taxon: Ontological transformation of NCBI Taxonomy

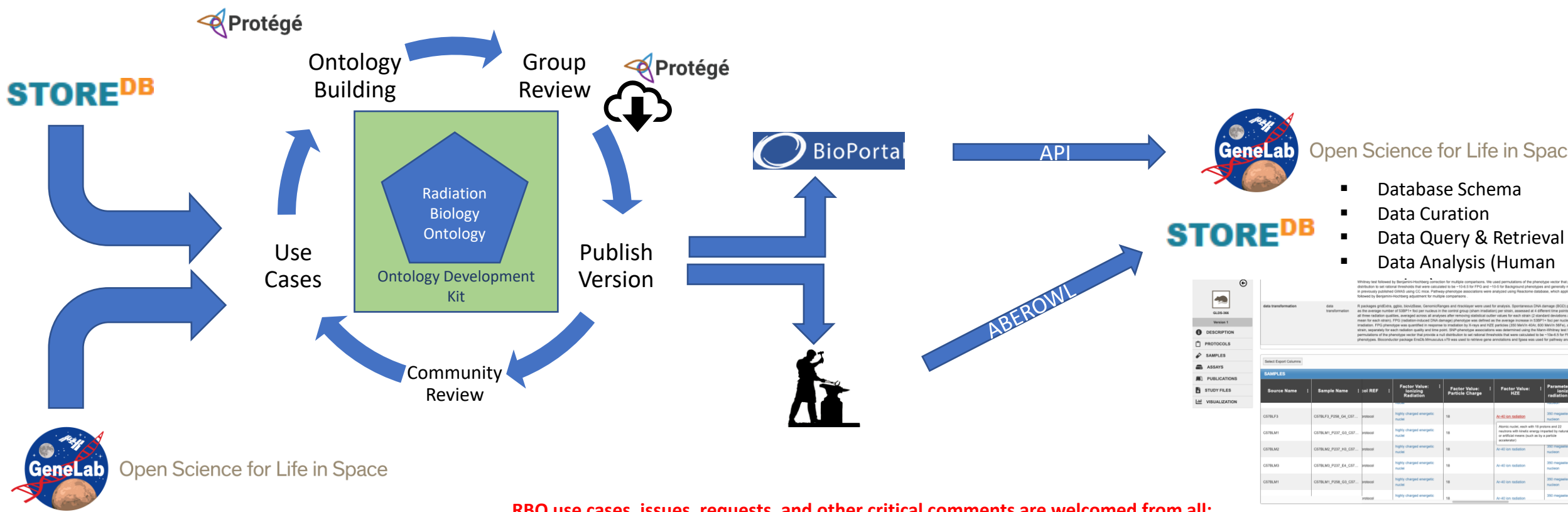
## Science Discovery Engine

- The Science Data Discovery Ontology was developed as part of NASA’s effort to create a Science Discovery Engine
- Both are being developed by the Science Mission Directorate TOPS group
- The SDE was built using the Sinequa application, a part of NASA EDP, and specifications for indexing data leveraging the SDDO



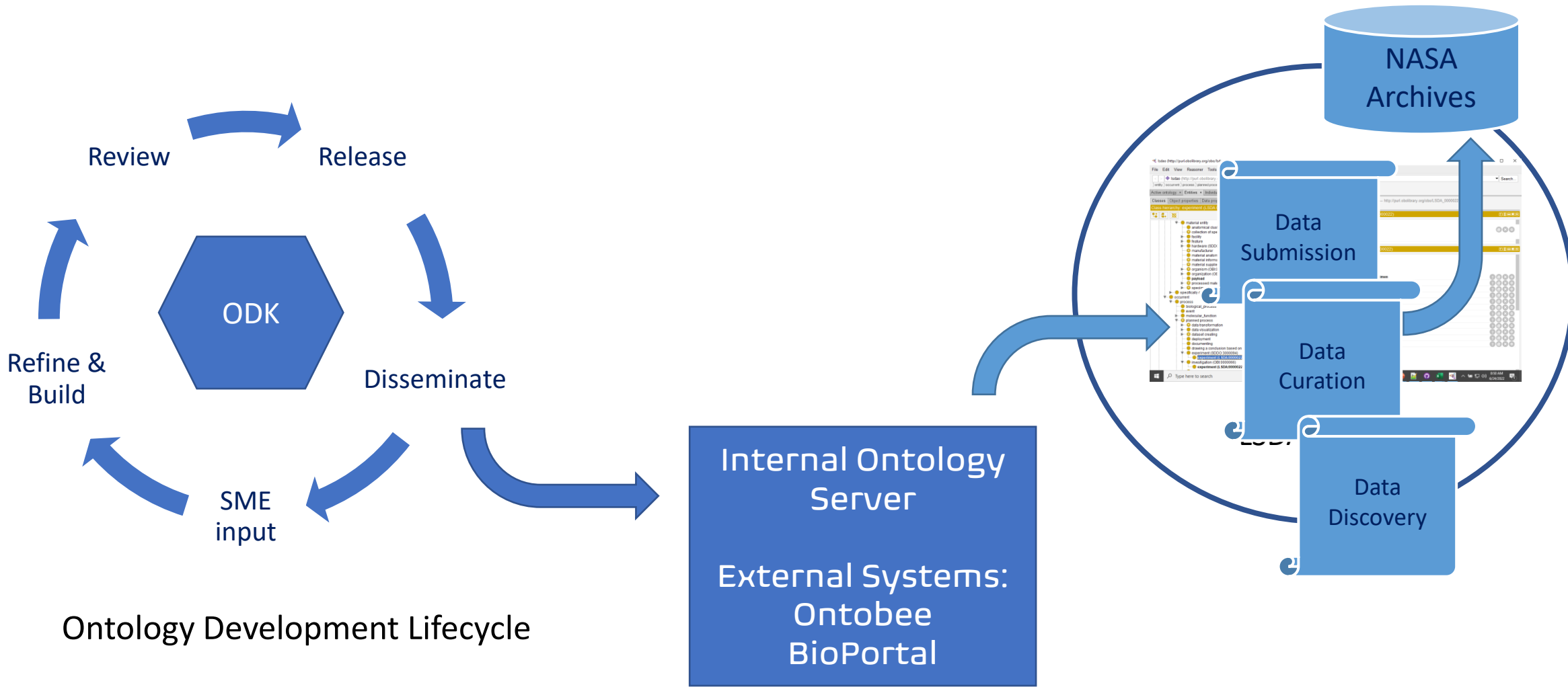
### Science Data Discovery Ontology





• An OBO Foundry domain ontology







# Summary



- Efforts towards frameworks that support semantic harmonization and data linkage increase transparency of science and FAIR compliance
- NASA life sciences repositories are working with the scientific research communities to develop and use knowledge resources such as
  - Metadata frameworks/models (e.g., ISATab)
  - Standard Vocabularies (like those that are part of OBO Foundry ontologies)
  - Citation and Licensing standards and services
- NASA is also developing FAIR compliance assessment and monitoring tools for these systems



Semantic Resources	Technologies	Tools
OBO Foundry ontologies	RDF/OWL	Protégé/WebProtege
W3C Ontologies		Ontology Dev Kit (INCATools) Gmake, OBO Library Robot
Dublin Core Ontologies		Cmap Tools  ELK, Hermit Reasoners